DOCUMENT RESUME

ED 226 045 TM 830 080

AUTHOR	Fuchs, Lynn S.; And Others
TITLE	Effects of Frequent Curriculum-Based Measurement and Evaluation on Student Achievement and Knowledge of Performance: An Experimental Study.
INSTITUTION	Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY	Department of Education, Washington, DC.
REPORT NO	IRLD-RR-96
PUB DATE	Nov 82
CONTRACT	300-80-0622
NOTE	65p.
AVAILABLE FROM	Editor, IRLD, 350 Elliott Hall, 750 East River Road, University of Minneapolis, MN 55455 ($3.00)
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Curriculum Development; *Disabilities; Elementary Education; *Evaluation Methods; Federal Legislation; Individualized Education Programs; Pretests Posttests; Program Development; Program Improvement; Rating Scales; *Self Concept; Special Education Teachers
IDENTIFIERS	*Stanford Diagnostic Reading Test

ABSTRACT

In special education, a merger between assessment and instruction is mandated by Federal law (PL 94-142), wherein teachers of handicapped pupils are required to specify individualized educational programs that include procedures for assessing students' progress toward goal attainment. The purpose of this study was to determine the effects of technically adequate, repeated curriculum-based measurement and evaluation procedures on students' reading achievement and on their knowledge concerning their own learning. Thirty-nine teachers, each having three or four students in the study, were assigned randomly to measurement/evaluation groups. Experimental group teachers employed continuous evaluation procedures while contrast group teachers employed conventional special education measurement and evaluation procedures. Students were pre- and posttested on an oral passage reading test and they were posttested on the Stanford Diagnostic Reading Test, Structural Analysis and Reading Comprehension subtests. Additionally, students were interviewed at the end of the study to assess their knowledge about their own progress. Analyses revealed that experimental students achieved more than contrast group students and were more knowledgeable about their learning. Results suggest that technically adequate, repeated curriculum-based measurement, when used by teachers to evaluate and modify programs, positively affects student achievement. Rating scales, an interview form and training schedules are appended. (Author/PN)

ED226045

University of Minnesota

Research Report No. 96

EFFECTS OF FREQUENT CURRICULUM-BASED MEASUREMENT AND EVALUATION ON STUDENT ACHIEVEMENT AND KNOWLEDGE OF PERFORMANCE: AN EXPERIMENTAL STUDY

Lynn S. Fuchs, Stanley L. Deno, and Phyllis K. Mirkin

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM EC
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

IRLD

Institute for Research on Learning Disabilities

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J Yss eldyke
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 830 060

IRLD

Director: James E. Ysseldyke

The Institute for Research on Learning Disabilities is supported by a contract (300-80-0622) with the Office of Special Education, Department of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled students.

During 1980-1983, Institute research focuses on four major areas:

- Referral
- Identification/Classification
- Intervention Planning and Progress Evaluation
- Outcome Evaluation

Additional information on the Institute's research objectives and activities may be obtained by writing to the Editor at the Institute (see Publications list for address).

The research reported herein was conducted under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent the official position of the Office of Special Education.

Research Report No. 96

EFFECTS OF FREQUENT CURRICULUM-BASED MEASUREMENT AND EVALUATION ON STUDENT ACHIEVEMENT AND KNOWLEDGE OF PERFORMANCE: AN EXPERIMENTAL STUDY

Lynn S. Fuchs, Stanley L. Deno, and Phyllis K. Mirkin

Institute for Research on Learning Disabilities

University of Minnesota

November, 1982

Abstract

The purpose of this study was to determine the effects of technically adequate, repeated curriculum-based measurement and evaluation procedures on students' reading achievement and on their knowledge concerning their own learning. Thirty-nine teachers, each having three to four students in the study, were assigned randomly to measurement/evaluation groups. Experimental group teachers employed continuous evaluation procedures while contrast group teachers employed conventional special education measurement and evaluation procedures. Students were pre- and posttested on an oral passage reading test and they were posttested on the Stanford Diagnostic Reading Test, Structural Analysis and Reading Comprehension subtests. Additionally, students were interviewed at the end of the study to assess their knowledge about their own progress. Analyses revealed that experimental students achieved more than contrast group students and were more knowledgeable about their learning. Implications for special education programming and assessment are discussed.

Effects of Frequent Curriculum-based Measurement and Evaluation on Student Achievement and Knowledge of Performance: An Experimental Study

The learning principles of educational psychology (Crow & Crow, 1963; Farnham-Diggory, 1972; Gagne, 1964) provide a theoretical framework for integrating measurement and evaluation with instruction. Properly conducted assessment provides at least two types of data that may be useful in enhancing achievement. First, assessment can help teachers better structure curriculum and teaching strategies to address students' individual needs and rates of progress. Second, measurement and evaluation can constitute an important source of feedback to a learner and can help the learner to see graded series of accomplishments and to recognize strategies for future goal attainment.

In special education, a merger between assessment and instruction is mandated by Federal law (PL 94-142), wherein teachers of handicapped pupils are required to specify Individualized Educational Programs (IEPs) that include procedures for assessing students' progress toward goal attainment. Substantive compliance with this law dictates that both measuring and evaluating student progress be on-going so that the adequacy of individual programs can be monitored continuously and improved as required (Deno & Mirkin, 1979). A small body of research (Beck, 1981; Haring, Maddux, & Krug, 1972; Mirkin, Deno, Tindal, & Kuehnle, 1980) has begun to accumulate that supports the hypothesis that such on-going measurement and evaluation may, in fact, facilitate student achievement. In the following discussion, the effects of frequent testing and the effects of on-going evaluation

are reviewed separately. Then, unified systems for integrating testing and evaluation with instruction are discussed. Finally, the purpose of the current study is presented.

Effects of Frequent Testing

Frequent testing appears to affect student achievement. One of the earliest published studies on the relation between achievement and testing was conducted by Gates (1917), who found that retention of new information was greatest when the presentation of each information unit was followed by a test. Gates' research set the stage for a series of studies investigating this notion. These studies (Jones, 1923; Keys, 1934; Noll, 1939) tended to support Gates' original findings. Early research on the positive relation between student learning and frequent testing has been corroborated by more recent studies. Robin (1976) found that frequent testing accounted for more variance in student achievement than any other variable in Keller's (1968) Personalized System of Instruction. Similarly, Omelich and Covington (1981) demonstrated with a path analysis that frequent testing was the factor that contributed most to pupil learning in the behavioral paradigm of direct, frequent assessment.

Frequent testing appears to enhance not only achievement but also student motivation. Feldhusen (1964) studied pupils who were administered 14 weekly quizzes that counted as one-fourth of their class grades. At least 80% of these students reported that frequent quizzes caused more learning, motivated them to study more than usual, and were effective in checking self progress. Of those students who reported that quizzes caused anxiety, 87% approved of that anxiety

because they believed that it helped them to learn.

Frequent testing also improves student motivation by providing feedback that can generate confidence in students regarding their ability to learn (Peckham & Roe, 1977). It can enhance an individual's perception of self-efficacy that can, in turn, influence thought patterns, actions, and emotional arousal that lead to goal attainment (Bandura, 1982; Prentice, 1961). Therefore, research supports the notion that frequent testing is related to students' growth, their attitude and approach to learning.

Effects of Continuous Evaluation

Unfortunately, it appears that frequent testing may be an insufficient condition for enhanced student development. Evidence suggests that teachers who collect student performance data do not necessarily use those data to make instructional decisions or to provide feedback to students (Baldwin, 1976; White, 1974). Additionally, when teachers fail to evaluate student test data continuously and systematically, learning is not improved (Tindal, Fuchs, Christenson, Mirkin, & Deno, 1981).

Attempts to Integrate Frequent Testing and Continuous Evaluation into Instruction

To increase the probability that students and teachers will use frequent testing results to influence student growth, educational psychologists have developed comprehensive systems for repeated testing and systematic evaluation. Keller (1968), for example, developed the Personalized System of Instruction (PSI), where individuals (a) progress at their own paces, (b) are tested repeatedly

until unit perfection is attained before being permitted to advance to new material, and (c) interact primarily with proctors who administer tests, provide instruction, and deliver immediate feedback.

Bloom (1971) has developed the model of mastery learning, which is based on Carroll's (1963) assumptions that learning can be broken down into a series of small tasks and that all students will master a given task provided they are given adequate time and instruction. The procedures for this model designate that consistent feedback be provided to students and teachers, and that instructional correctives also be provided as required. As with Keller's PSI, frequent testing and feedback are integral parts of this learning model.

Other educational psychologists have built upon previous work by incorporating into Keller and Bloom's basic principles the operant methodology of repeated behavior sampling and time-series analysis. Lovitt (1977), Lindsley (1971), and White and Haring (1980) have developed models of direct, frequent testing and evaluation where a teacher collects repeated short samples of a student's behavior within a curriculum, over a time period, and under different teaching strategies. The teacher then applies the methods of time-series analysis to the data in order to determine the effectiveness of specific program dimensions.

Unfortunately, all of the curriculum-based testing and evaluation models referenced above require teachers continuously to create their own testing materials; little is known regarding the technical characteristics of such self-prepared measurement and testing procedures. As Popham (1980) illustrates, creating a valid, reliable

criterion-referenced test is laborious, difficult, and time consuming. Additionally, since variations in testing procedures have a dramatic bearing on the adequacy of norm-referenced standardized tests (Sattler, 1974), variations in procedures for repeatedly testing student performance on curriculum objectives also may affect the technical adequacy of the measurement.

Therefore, while the development of curriculum-based testing and evaluation procedures may address the need to integrate measurement and evaluation with instruction, some models may have neglected the need for edumetrically adequate testing and evaluation instruments and procedures. Such approaches to measurement may create unreliable and invalid data bases with which to formulate educational decisions.

In response to the need for technically adequate measurement and evaluation, Mirkin, Deno, Fuchs, Wesson, Tindal, Marston, and Kuehnle (1981) developed a model of curriculum-based measurement and time-series evaluation that includes five decision-making steps. At each step, the procedures (described below) demonstrate edumetric adequacy and logistical feasibility (Fuchs, Mirkin, Deno, Marston, & Tindal, 1982).

> Step 1: Selection of the program objective. The program objective specifies the date and the segment of the curriculum on which the student will perform, and the specific level of proficiency. So, a student's reading objective might specify: By June 2, when presented with stories from page 29 to page 307 in Level 11 for one minute, George will read aloud at least 70 words correctly with 7 or fewer errors. In spelling, a student's objective might read: By February 15, when dictated words from Level 9 for two minutes, Henry will write at least 10 words correctly with no more than 2 errors.
>
> Step 2: Selection of the measurement domain. The measurement domain is the set of material from which test

samples are drawn. Once an objective has been established, the measurement domain is defined as that level in the curriculum where the student is expected to be proficient. Consequently, given the above reading objective, the measurement domain would be reading passages from pages 29 to 307 in Level 11. Given the above spelling objective, the measurement domain would be the words from Level 9.

Step 3: Selection of the measurement task and procedure. The measurement task is a randomly selected portion of the measurement domain. For example, using the objectives cited above, a passage randomly drawn from pages 29 to 307 would be the measurement task in reading; a set of words randomly drawn from Level 9 would be the measurement task in spelling. Each time measurement occurs the stimulus material changes; a new sample from the domain is drawn randomly. The measurement procedure varies across academic areas but, within a given area, the procedure is standardized across days. (See Mirkin et al., 1981, for recommended procedures in reading, spelling, written expression, and social behavior.) Student performance is measured at least three times per week; measurement occurs daily if possible.

Step 4: Selection of a time-series data display. Student performance data are displayed on a Cartesian graph, with "school days" labeled on the horizontal axis and "behavior" on the vertical axis. A data point represents the student's performance level on the measurement domain for a given day. An example of a graph is provided in Figure 1. In this instance, the graph represents the "Number of Words Read Aloud from Text Passages in Level 2 Per Day." Increases in the data points across the measurement days indicate reading improvement and progress toward the student's objective. Each solid vertical line of the graph represents the introduction of a program change.

Insert Figure 1 about here

Step 5: Selection of data evaluation procedures. Two alternatives to data analysis exist, experimental and therapeutic analysis. In experimental analysis, program changes are introduced regularly and are treated as experimental hypotheses concerning their effect on a student's performance. The methods of time-series analysis (Sidman, 1960) are employed to summarize and interpret student performance data. Within program phases, data are summarized with a split-median solution trend line (White,

1971), an index of variability such as total bounce (Pennypacker, Koenig, & Lindsley, 1972), and an index of step-up or step-down. Data interpretation is relative; effectiveness of a program phase is evaluated by comparing descriptive statistics across treatments. Therefore, changes in median, level, trend, and/or variability between adjacent phases are inspected and interpreted to formulate decisions about the effectiveness of a specific program change.

In therapeutic analysis, the student's objective is marked on the graph with an X at the intersection of the desired performance level and the expected attainment date. Then, a line of required progress, connecting the student's baseline median score with the desired level is drawn on the graph. This line is called the dynamic aimline. Data interpretation involves the application of the following rule: If 3 consecutive data points fall below the trend line, then the program is judged ineffective and a change in the program is introduced.

Evidence (Fuchs, Mirkin, Deno, Marston, & Tindal, 1982) suggests that a combination of the two data-utilization methods, which borrows more heavily from the therapeutic method, may be most useful and feasible. In this approach, recommended by Mirkin et al. (1981), teachers draw a dynamic aimline on the graph. Then, a split-median trend line on 7 to 10 student performance data points is graphed and compared to the slope of the dynamic aimline. If the student performance slope is less steep than the aimline, a program change is introduced.

With this model, technically adequate measurement and evaluation are integrated with instruction. Program development is cumulative, data based, and on-going. Time-series displays of student performance data are inspected; evaluations concerning the effectiveness of individual programs are made; successful program changes are maintained and incorporated into new program phases, while unsuccessful changes are terminated. In this way, technically adequate, curriculum-based time-series measurement and evaluation may provide an adequate data base with which educators can evaluate and improve individual programs formatively, to increase the likelihood that student objectives will be realized.

Purpose of the Study

The purpose of the study described here was to determine the educational effects of the technically adequate, repeated curriculum-

based measurement and evaluation procedures developed by Mirkin et al. (1981). Specifically, in special education programs of a large metropolitan school district, the effects of such measurement and evaluation on students' reading achievement and their knowledge about their own learning were explored.

Method

Subjects

Subjects were 39 New York City public school teachers who volunteered to participate in the project. From their caseloads, teachers each selected three or four students for inclusion in the study; then, within schools, each teacher was assigned randomly to either an experimental or a contrast group. Four teachers in each group had participated in a pilot program during the previous academic year.

In the experimental group, teachers (3 male, 15 female) had taught special education for an average 3.79 years (SD = 2.85). Students in the experimental group (51 male, 13 female) read an average 3.48 years below grade level (SD = 1.87). Their age appropriate grade averaged 5.79 (SD = 1.66); 49% were placed in programs for emotionally handicapped students, 32% in programs for brain-injured students, and 19% in resource programs.

Contrast group teachers (2 male, 19 female) had taught handicapped children for an average 3.59 years (SD = 2.72). Students in the contrast group (57 male, 20 female) read an average 2.35 years below grade level (SD = 1.94). Their age appropriate grade level averaged 5.45 (SD = 1.65); 51% were placed in programs for emotionally

handicapped students, 30% in resource programs, and 5% equally distributed across programs for brain-injured, physically-handicapped, and educable mentally retarded children.

Measures

Accuracy of implementation rating scale (AIRS). The AIRS (Wesson, Deno, Maruyama, Sevcik, Skiba, King, & Mirkin, 1982; see Appendix A) was employed to measure the accuracy with which the experimental procedures (Mirkin et al., 1981) were implemented. The first two items of the 13-item rating instrument, Administering the Measurement Task and Selecting the Stimulus Material, are completed following direct observation of student-teacher interaction. The next five items, Sampling for Instructional Level, Baseline, Graphing Format, Aimline, and Timing of Instructional Changes, are completed after inspecting the student graph (see Appendix B). The Long-range Goal, Short-term Objective, and Measurement System items are scored after inspecting the Goal and Objective Form (see Appendix C). The Instructional Plan, Substantial Changes, and Clear Changes scores are based on inspection of the Instructional Change Form (see Appendix D). For descriptive purposes, eight items were included in one group called Measurement since they all related to establishing and conducting measurement (cf. Wesson et al., 1982).

Student interview. A four-item interview schedule was designed to assess students' knowledge of (a) their reading progress, (b) their reading goals, and (c) the likelihood that they would attain their reading goals during the current academic year. The last two items required interviewers to assess the accuracy of student responses

against graphs and records. (See End-of-year Student Questionnaire in Appendix E.)

Passage reading test. A series of three reading passages from a third grade book of the Ginn 720 reading series (Ginn & Company, 1976) was employed in measurement. Two passages were sampled randomly from the text and one passage was chosen to represent the readability level of the last 25% of the text. (See Fuchs & Deno, 1981, for the selection procedure.)

Stanford Diagnostic Reading Test. Two subtests, Structural Analysis and Reading Comprehension, of the Stanford Diagnostic Reading Test (Karlsen, Madden, & Gardner, 1976), Green Level, Form A, were employed as measures. The Structural Analysis subtest (SA) measures a child's decoding skills through the analysis of word parts. The SA asks pupils to (a) identify the first syllable of two-syllable words, and (b) find meaningful word parts and blend those parts into real words. Internal consistency reliability for the SA, Form A, was .93 for grade 3 students, and .95 for grades 4 and 5 children. Criterion-related validity with respect to the reading tests of the Stanford Achievement Test ranged from .62 to .85, with a correlation coefficient of .85 between the SA and the Total Reading Score on the Stanford Reading Tests.

The Reading Comprehension subtest (RC) assesses literal and inferential comprehension through short reading passages presented in a multiple choice format and through short passages followed by questions. The passages represent different subject matter areas at a third through fifth grade readability level. Internal consistency

reliability was reported at .96 for the RC, Form A, at grades 3 through 5. Criterion validity with respect to the reading tests of the Stanford Achievement Test was reported, with correlation coefficients ranging from .68 to .90.

Procedure

Pretest. During November, students were pretested on the third grade Ginn 720 reading passages, with one passage administered each week during three successive weeks. Each student was tested individually by one of three teacher trainers who were skilled in the reading passage administration procedure (see Mirkin et al., 1981). In this reading passage test, pupils read orally for one minute; words correct and errors per minute were scored.

Treatments. The experimental treatment is described in Procedures to Develop and Monitor Progress on IEP Goals (Mirkin et al., 1981). Employing these procedures in the area of reading, the experimental group teachers first wrote curriculum-based IEP goals and objectives. The annual goal specified the segment of the curriculum and the date on which a student would read at a certain rate and accuracy. The objective contained supplementary information; it indicated the weekly rate at which the student would improve in order to meet the annual goal (see Goal and Objective form in Appendix C). After setting goals and objectives, teachers developed curriculum-based measurement systems to match specific goals and objectives. Then, they were to measure students at least twice weekly and utilize those data to determine when to introduce program changes to increase the probability that students would achieve their goals. By January

4, all teachers had written goals and objectives, and were measuring and graphing student performance. By February 10, all teachers were employing data utilization rules to determine when to make instructional changes (cf. Mirkin et al., 1981).

A sample Goal and Objective Form, Graph, and Instructional Change Form for Michael appear in Figures 2, 3, and 4, respectively. The Goal and Objective Form states that, in the 19 weeks remaining in the school year, Michael will improve his reading in Level 2 of the SRA passages so that he reads 85 words correctly per minute with no more than 8 errors (see Figure 2). This annual goal represents approximately a 2.5 times improvement over Michael's baseline rate of 35 words per minute (see Figure 3). This annual goal then was divided into 19 weekly improvement rates; Michael's objective states that he will improve at an average increase of 2.6 words correct each week with no increase in errors.

Insert Figures 2-4 about here

Michael's graph (see Figure 3) represents his words read per minute each school day in Level 2 of the SRA passages. The first three days of data depict Michael's baseline performance; the diagonal line connecting the baseline median performance with the X is the dynamic aimline, which represents Michael's objective or the daily rate at which he must perform in order to meet the annual goal.

The vertical lines on Michael's graph signify the introduction of program changes, which are indicated by the letters B, C, and D. Each

program change is detailed on the Instructional Change Form (see Figure 4). At least two times each week, Michael read from a randomly selected passage from Level 2 for one minute; the number of words he read correctly and the number of errors were scored and graphed on Figure 3. Every 7 to 10 data points, Michael's teacher inspected the graph. If the slope calculated by the split median solution (White, 1971) of Michael's performance across the 7 to 10 days was less than the slope of the dynamic aimline, then the teacher introduced a program change. As Figure 3 illustrates, Michael's performance improved dramatically over his previous performance with the introduction of the third program change.

With the Goal and Objective Form, the Graph, and the Instructional Change Form, Michael's educational program and progress toward goals were evaluated formatively. In response to measurement data, Michael's program was modified throughout the treatment to improve the likelihood that Michael would achieve his annual goal.

In the contrast treatment, teachers measured and evaluated students' progress toward goals using conventional special education procedures; these typically include relatively infrequent teacher-made tests and informal observation (Fuchs, Deno, & Mirkin, 1982; Fuchs, Fuchs, & Warren, 1982).

Training. Each of three teacher trainers (TTs) was assigned to a set of schools and to the experimental and contrast teachers within that set of schools. TTs provided training to teachers during weekly meetings from November through May. During these meetings, TTs taught the treatment procedures to the experimental group teachers. They

taught the contrast teachers strategies for diagnosing and treating learning and behavior problems, for structuring and managing their instructional groups, and for using audio-visual equipment and paraprofessionals. The TTs spent similar amounts of time with both groups of teachers: Experimental group teachers received an average 23.52 hours of individual meeting time (SD = 5.95), whereas contrast group teachers met individually with the trainers for an average 20.60 hours (SD = 6.22). A *t* test on the difference between the hours of training received by the two groups revealed no statistically significant difference. (See Appendix F for sample experimental and contrast training schedules.)

Data collection. TTs were taught to measure experimental teachers' performance on the AIRS during one five-hour training session. In January, April, and late May, a TT observed a randomly selected student for each experimental group teacher while the edocator was preparing for and measuring the student's reading behavior. Immediately following an observation, the trainer scored items 1 and 2 of the AIRS. Then, the trainers collected the experimental students' Goal and Objective Forms, Graphs, and Instructional Change Forms, and gave them to IRLD staff who used the documents to score items 3 through 13 of the AIRS.

During June, students were tested in groups of 4 to 10 children on the Stanford Diagnostic Reading Test, SA and RC subtests. Additionally, the three third grade reading passages were administered individually in standard fashion (Mirkin et al., 1981) during one session. All testing was conducted by the TTs. In June, a subsample

of students was interviewed by the trainers on the End-of-the-year Student Interview Schedule.

Data Analysis

Because teachers, rather than students, were assigned randomly to treatments, student reading scores for each teacher on each dependent measure were averaged, and "teacher" was designated as the experimental unit of analysis. Comparisons of the pretest oral reading error scores of the two groups revealed a statistically significant difference favoring the contrast group, $t(37) = 2.03$, $p < .05$ (mean difference = 1.95). The reading posttest scores, therefore, were analyzed by means of a two-way multivariate analysis of covariance (MANCOVA) as well as follow-up two-way univariate analyses of covariance (ANCOVAs). The experimental factor was measurement/evaluation treatment and the blocking factor was teacher trainer; the pretest error variable was entered as the covariate in each analysis. Prior to using the analysis of covariance, the assumption of homogeneous regression coefficients was tested and scattergrams were inspected; assumptions of equal regression slopes and linearity of Y on X appeared tenable.

Responses on the End-of-the-year Student Interview were analyzed employing chi-square tests. Measurement/evaluation treatment and the relevant item on the survey were the two dimensions of each contingency table. The chi-square statistic was selected over correlations because of the nominal, dichotomous nature of the measurement/evaluation factor.

Results

The analysis of the results addressed three questions. First, to what degree did the teachers implement the measurement and evaluation system? Second, did the students of the experimental and contrast teachers differ in their reading achievement? Finally, did the students of the experimental and contrast teachers differ with respect to their knowledge of their reading goals and progress?

System Implementation

There were three indices of the extent to which, and the accuracy with which, the experimental teachers actually implemented the measurement and evaluation system: the AIRS, the weekly number of data points on students' graphs, and the number of changes teachers introduced into their students' programs.

Table 1 displays means and standard deviations for the three data collection times on the 13 items of the AIRS, the Measurement group of items, and the average across the Measurement group and the other five items (the overall total). For the overall total, ratings of the teachers on the 5-point scale were 3.32, 3.66, and 3.85, respectively, for the three data collection times.

Insert Table 1 about here

Teachers measured students' reading performance an average 2.14 times per week (SD = .44) during the first six weeks of the treatment. They measured students' reading performance an average 1.87 times per week (SD = .40) during the second six weeks, and an average 1.37 times

per week (SD = .77) during the last six weeks.

Over the 18-week treatment, teachers introduced a median two modifications in their students' programs. In Table 2, a frequency count of teachers introducing different numbers of changes is shown. The mean, standard deviation, and median number of changes introduced during each trimester of the study also are displayed. Teachers introduced an average .95, 1.01, and .22 changes into students' programs during each trimester, respectively.

Insert Table 2 about here

Student Achievement

The two-way multivariate analysis of covariance was conducted on the four posttest reading variables: the mean number of words correct per minute on the passage reading test, the mean number of errors per minute on the passage reading test, the raw score on the SA, and the raw score on the RC. Wilk's lambda criterion was used to test for equality of group centroids. The value calculated with the Wilk's lambda procedure was transformed into an F value through Rao's approximation. The test of lambda produced F values that were statistically significant for the measurement/evaluation factor, $F(4,29) = 3.80$, $p < .05$, and for the teacher trainer factor, $F(8,58) = 12.99$, $p < .001$. Therefore, univariate two-way ANCOVAs were computed.

These ANCOVAs revealed statistically significant differences between the measurement/evaluation conditions on three of four dependent variables. Table 3 displays, for each measurement

evaluation group and for each dependent variable, the mean, the standard deviation, the mean adjusted for the independent and covariate, and the F and p-values.

Insert Table 3 about here

The univariate ANCOVAs included teacher trainer as a second, blocking factor to increase the statistical power of the analysis rather than to explore an interesting, related research question. Therefore, although there were statistically significant differences between teacher trainer groups on two of four variables, those results are not presented further or discussed. The absence of any statistically significant interaction between the measurement evaluation and teacher trainer factors renders any additional discussion of the teacher trainer conditions extraneous to the purposes of this paper.

Students' Knowledge about their Learning

A final set of analyses was conducted to ascertain differences between experimental and contrast conditions with respect to student knowledge of their own goals and progress. Figure 5 is a display of the percentages of experimental and contrast group student responses when they were asked during the End-of-the-year Student Interview, "What kind of progress have you made in reading this year?" As one can see, the trend of responses for each group was similar. A chi-square analysis on the proportional responding of experimental and contrast groups to this question revealed no significant difference.

A visual analysis of Figure 5, however, reveals that the contrast group students' curve was skewed more positively than that of the experimental group students, with a greater percentage of control group students describing their progress as poor, and with a smaller percentage of contrast group children labeling their progress as great.

Insert Figure 5 about here

The remaining questions on the End-of-the-year Student Interview asked (a) whether students knew their reading goals and, if so, could they state them, (b) whether they would meet their reading goals this year, and (c) how they knew whether they would meet their goals. There was a statistically significant relation between measurement/evaluation treatment and whether students said they knew their goals, $\chi^2(1) = 4.17$, $p < .05$, with experimentals typically saying they knew their goals and with contrasts typically saying they did not know their goals. Further, there was a statistically significant relation between treatment group and students actually stating their goals, $\chi^2(1) = 3.91$, $p < .05$, with experimentals frequently stating their goals and with contrasts infrequently stating their goals.

There was no statistically significant relation between treatment and whether students stated that they thought they would meet their goals. Yet, the accuracy of these statements was related to treatment group, $\chi^2(1) = 6.61$, $p < .01$, with a greater percentage of experimental group statements correct. Additionally, when asked

how they knew whether they would meet their goals, experimental students tended to say that they relied on their graphed data whereas contrast students tended to say that they "just thought so," $\chi^2(1) = 9.47$, $p < .005$.

Discussion

The purpose of this study was to determine the educational effects of technically adequate, repeated curriculum-based measurement and evaluation procedures. Using these procedures, the experimental teachers were (a) to measure and graph students' proficiency in reading aloud from text passages at least twice weekly, and (b) to compare, at frequent intervals, the students' actual increases in proficiency with increments required to achieve the level of proficiency specified in the IEP goal. When students' actual progress trends compared unfavorably with their required increases, teachers were to introduce a new dimension into the students' programs in an attempt to stimulate better student progress. Comparison teachers employed typical special education measurement and evaluation procedures.

The findings indicated that experimental teachers implemented their treatment moderately well for most of the treatment period. Although teachers conformed to the prescribed measurement schedule fairly well during the first 12 weeks of the study, their rate of measuring fell during the last six weeks of the treatment. Analogously, the mean number of changes introduced by teachers into students' programs dropped sharply during the last trimester of the study. These findings corroborated prior research (Tindal et al.,

1981), which suggested that teachers tend to measure students less frequently and evaluate data less responsively than prescribed. Nevertheless, the extent of teachers' implementation was adequate during most of this study, and their accuracy in implementing experimental procedures remained fairly stable across the entire treatment period.

The results indicate that implementation of the frequent, direct measurement and evaluation procedures affected positively both student achievement and student awareness of their own achievement. Students of experimental teachers performed better than students of comparison teachers on virtually all achievement measures, rate and accuracy in reading aloud from text materials, and the Structural Analysis and Reading Comprehension subtests of the Stanford Diagnostic Reading Test. It is worth noting that while the former measure (reading aloud from text) was used throughout the study, the latter two were not directly measured by the teachers as part of the evaluation system. The findings suggest that when teachers repeatedly employ the simple one-minute test of reading aloud from text passages to index student progress, they can interpret student gains as representing general reading achievement; that is, the data apparently validly reflect fluency, decoding, and comprehension. The only measure on which experimental and contrast group performance was undifferentiated was error scores on the third grade oral reading passages. This may be explained by the poor reliability of such error scores (Deno, Mirkin, Chiang, & Lowry, 1980; Fuchs & Deno, 1981; Fuchs, Deno, & Marston, 1982). Consequently, the results of this study suggest that

technically adequate, repeated curriculum-based measurement, when used by teachers to evaluate and modify programs, positively affects student achievement.

In addition to achieving better, the students in this study were more knowledgeable about their own learning when their teachers used systematic measurement and evaluation procedures. As compared to pupils whose progress was measured and evaluated via conventional special education practice, students who were measured and evaluated repeatedly and systematically in their curricula (a) more frequently said they knew their goals, (b) more often actually stated their goals, (c) were more accurate in their estimates of whether they would meet their goals, and (d) more typically reported that they relied on data to formulate estimates of whether they would meet their goals. The results revealed that repeated, direct systematic measurement and evaluation enhances not only students' reading achievement, but also their knowledge concerning their own learning. These outcomes are theoretically and socially important. On the one hand, they support the hypothesis of many educational psychologists (Bandura, 1982; Crow & Crow, 1963; Farnham-Diggory, 1972; Prentice, 1961), that students' knowledge of their learning may improve academic performance. On the other hand, increased participation by students in their own education, itself, is often regarded as an important educational goal.

This study generally provides support for increased use of systematic, on-going measurement and evaluation of student progress by teachers. It provides evidence that individual special education programs can be monitored continuously and improved as required to

increase the likelihood of student gains. The findings of this research contradict the conventional argument that teachers do not need to use frequent measurement and evaluation because (a) they are already sufficiently aware of student achievement, and (b) such procedures are inefficient. Teachers who used frequent measurement and systematic evaluation were more effective in enhancing student growth and student awareness of their educational programs.

References

Baldwin, V. Curriculum concerns. In M. A. Thomas (Ed.), Hey, don't forget about me. Reston, VA: Council for Exceptional Children, 1976.

Bandura, A. Self-efficacy mechanism in human agency. American Psychologist, 1982, 37(2), 122-147.

Beck, R. Curriculum management through a data base. Great Falls, MT: Great Falls Public Schools, 1981.

Bloom, B. S. Mastery learning. In J. H. Block (Ed.), Mastery learning: Theory and practice. New York: Holt, Rinehart, & Winston, 1971.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Crow, L. D., & Crow, A. Readings in human learning. New York: David McKay, 1963.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1979. (ERIC Document Reproduction Service No. ED 185 770)

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980. (ERIC Document Reproduction Service No. ED 197 507)

Farnham-Diggory, S. Cognitive processes in education: A psychological preparation for teaching and curriculum development. New York: Harper & Row, 1972.

Feldhusen, J. F. Student perceptions of frequent quizzes and post-mortem discussion of tests. Journal of Educational Measurement, 1964, 1, 51-54.

Fuchs, L. S., & Deno, S. L. The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981. (ERIC Document Reproduction Service No. ED 212 662)

Fuchs, L. S., & Deno, S. L., & Marston, D. Use of aggregation to improve the reliability of simple direct measures of academic performance (Research Report No. 94). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. The effect of continuous evaluation procedures on special education teacher decision making (Research Report No. 97). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Fuchs, L. S., Mirkin, P. K., Deno, S. L., Marston, D., & Tindal, G. Considerations for designing a continuous evaluation system: An integrative review (Monograph, in preparation). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

Gagne, R. M. The implications of instructional objectives for learning. In C. M. Lindvall (Ed.), Defining educational objectives. Pittsburgh: University of Pittsburgh, 1964.

Gates, A. Recitation as a factor in memorizing. Archives of Psychology, 1917, 40, 1-104.

Ginn and Company. Reading 720. Lexington, MA: Ginn (Xerox Corp.), 1976.

Haring, N., Maddux, L., & Krug, D. A. Investigation of systematic instructional procedures to facilitate academic achievement in mentally retarded disadvantaged children (Final Report). Seattle: University of Washington, Experimental Education Unit, 1972.

Jones, H. E. Experimental studies of college teaching. Archives of Psychology, 1923, 68, 5-70.

Karlsen, B., Madden, R., & Gardner, E. F. Stanford diagnostic reading test. New York: Harcourt Brace Jovanovich, 1976.

Keller, F. S. Goodbye, teacher.... Journal of Behavior Analysis, 1968, 1, 79-89.

Keys, N. The influence on learning and retention of monthly tests. Journal of Educational Psychology, 1934, 25, 427-436.

Lindsley, O. R. Precision teaching in perspective: An interview with Ogden R. Lindsley. Teaching Exceptional Children, 1971, 3(3), 114-119.

Lovitt, T. In spite of my resistance, I've learned from children. Columbus, OH: Charles E. Merrill, 1977.

Mirkin, P. K., Deno, S. L., Fuchs, L. S., Wesson, C., Tindal, G., Marston, D., & Kuehnle, K. Procedures to develop and monitor progress toward IEP goals. Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1980. (ERIC Document Reproduction Service No. ED 197 510)

Noll, V. H. The effect of written tests upon achievement in college classes: An experiment and a summary of evidence. Journal of Educational Research, 1939, 32, 351-357.

Omelich, C. L., & Covington, M. V. Why behavioral instruction works? A components analysis of performance and motivational outcomes. Paper presented at the meeting of the American Educational Research Association, Los Angeles, April, 1981.

Peckham, P. D., & Roe, M. D. The effects of frequent testing. Journal of Research and Development in Education, 1977, 10(3), 40-50.

Pennypacker, H. S., Koenig, C. H., & Lindsley, O. R. Handbook of the standard behavior chart (prelim. ed.). Kansas City, KS: Precision Media, 1972.

Popham, W. J. Domain specification strategies. In R. A. Beck (Ed.), Criterion-referenced measurement: The state of the art. Baltimore: Johns Hopkins University Press, 1980.

Prentice, W. C. H. Some cognitive aspects of motivation. American Psychologist, 1961, 16, 503-511.

Robin, A. L. Behavioral instruction in the college classroom. Review of Educational Research, 1976, 46, 313-354.

Sattler, J. M. Assessment of children's intelligence. Philadelphia: W. B. Saunders, 1974.

Sidman, M. Scientific research: Evaluating experimental data in psychology. New York: Basic Books, 1960.

Tindal, G., Fuchs, L. S., Christenson, S., Mirkin, P. K., & Deno, S. L. The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1981.

Wesson, C., Deno, S., Maruyama, G., Sevcik, B., Skiba, R., King, R., & Mirkin, P. The relationships among implementation of an on-going evaluation system, structure of instructional programs, and achievement (Research Report, in preparation). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities, 1982.

White, O. R. A pragmatic approach to the description of progress in the single case. Unpublished doctoral dissertation, University of Oregon, 1971.

White, O. R. Evaluating education process (Working paper). Seattle: University of Washington, Child Development and Mental Retardation Center, Experimental Education Unit, 1974.

White, O. R., & Haring, N. G. Exceptional teaching (2nd ed.). Columbus, OH: Charles E. Merrill, 1980.

Table 1

Means and Standard Deviations on the Accuracy of Implementation Rating Scale (AIRS) During Three Data Collection Times[a]

Items	Time 1		Time 2		Time 3	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Administration of Task	4.78	.43	4.83	.38	4.78	.64
Sampling Instructional Level	3.11	1.32	3.28	1.41	3.28	1.41
Taking Baseline	3.56	1.20	3.50	1.15	3.50	1.15
Graphing	4.28	.57	4.17	.51	4.17	.51
Drawing Aimline	4.44	1.29	4.44	1.29	4.44	1.29
Writing Goal	4.94	.24	4.94	.24	4.94	.24
Writing Objective	4.78	.94	4.78	.94	4.78	.94
Describing Measurement System	5.00	.00	5.00	.00	4.89	.47
Measurement[b]	4.30	.39	4.36	.38	4.28	.33
Writing Instructional Plan	2.11	1.32	2.67	1.28	2.64 (N=11)	1.36
Creating Substantial Changes	3.31 (N=13)	.75	3.94 (N=16)	.85	3.00 (N=2)	2.82
Describing Clear Changes	2.77 (N=13)	1.17	2.88 (N=16)	1.20	5.00 (N=2)	.00
Timing Instructional Changes	2.61	1.58	3.83	.92	3.35 (N=17)	1.69
Selecting Stimulus Material	4.61	.70	4.91	.24	4.89	.64
Overall Total[c]	3.32	.62	3.66	.52	3.85	.68

[a] N=18 teachers, except as indicated in parentheses.

[b] The measurement group combined eight items related to establishing the measurement system and conducting measurement.

[c] The overall total was based upon the addition of the measurement group and each of the remaining five items.

Table 2

Number of Changes Introduced into Experimental Students' Programs During Each Third of the Treatment Period

	First Third	Second Third	Third Third
Frequency of 0 Changes	21	15	53
Frequency of 1 Change	26	35	6
Frequency of 2 Changes	14	10	4
Frequency of 3 Changes	2	3	0
Mean Changes	.95	1.01	.22
Standard Deviation	.83	.77	.55
Median Changes	1	1	0

Table 3

Means, Standard Deviations, Adjusted Means, and ANCOVA Results on Four Dependent Variables

	Measurement/Evaluation Treatment							
	Experimental			Contrast			F test	
	$\bar{X}$	SD	Adjusted $\bar{X}$	$\bar{X}$	SD	Adjusted $\bar{X}$	F(1,32)	p
Words correct per minute	70.23	45.75	78.93	51.29	38.18	43.81	4.221	.048
Errors per minute	5.63	2.08	5.21	5.64	2.75	5.99	.001	.982
SA	39.79	12.28	40.74	29.65	15.34	28.84	7.194	.011
RC	43.95	10.52	45.37	33.02	15.39	31.81	4.222	.048

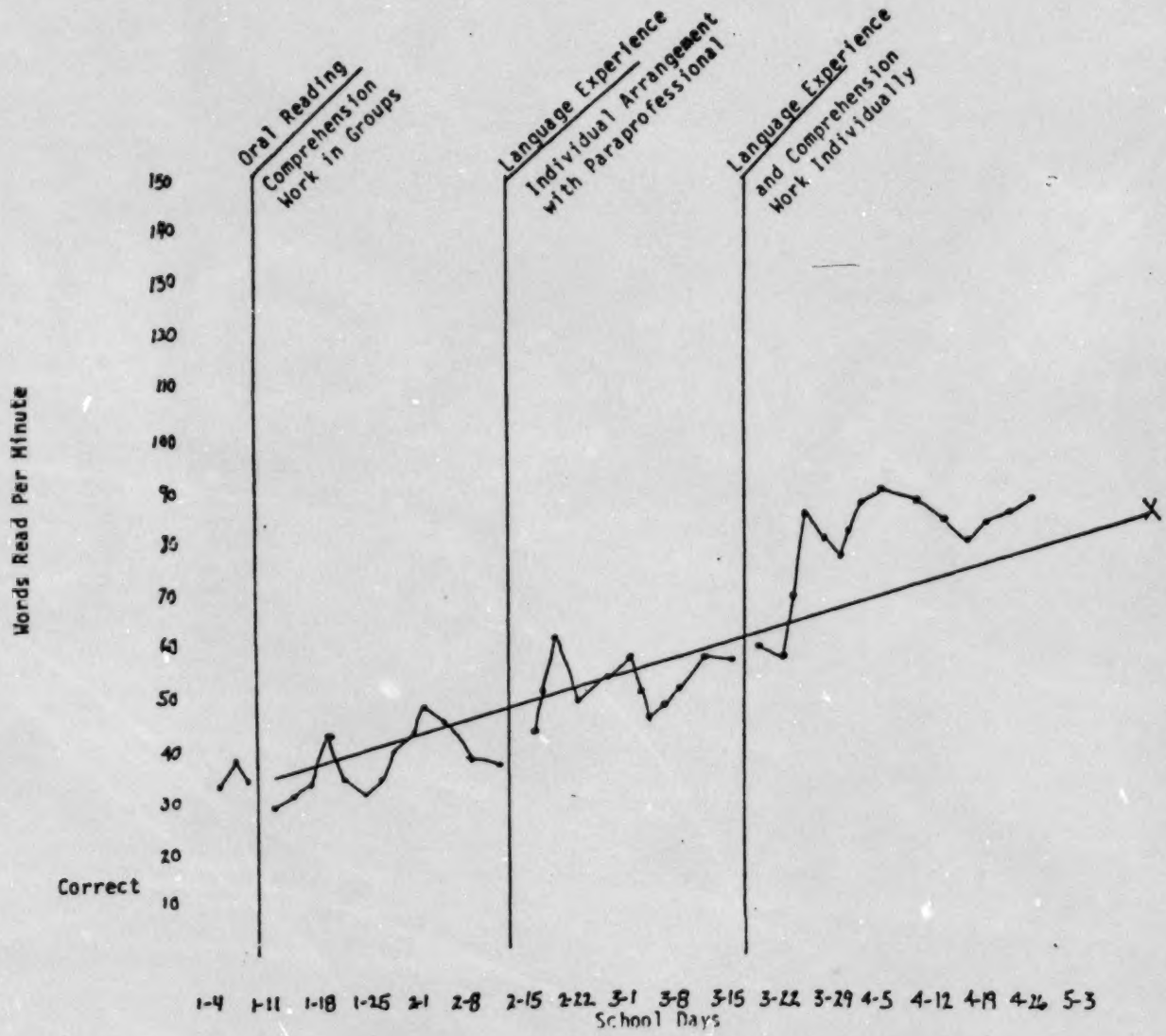


Figure 1. Example of a graph of student performance data.

GOAL In 19 weeks (# school weeks until year's end), when provided with stories from grade level 2 - SRA passages (Level #, series), Michael (student's name) will read aloud at the rate of 85 (wpm correct), with no more than 8 (#) errors.

OBJECTIVE Each successive week, when presented with a random selection from Grade level 2 - SRA passages (same level # and series as above), the student will read aloud at an average increase of 2.6 (#) wpm and no increase in errors.

Figure 2. Michael's Goal and Objective Form

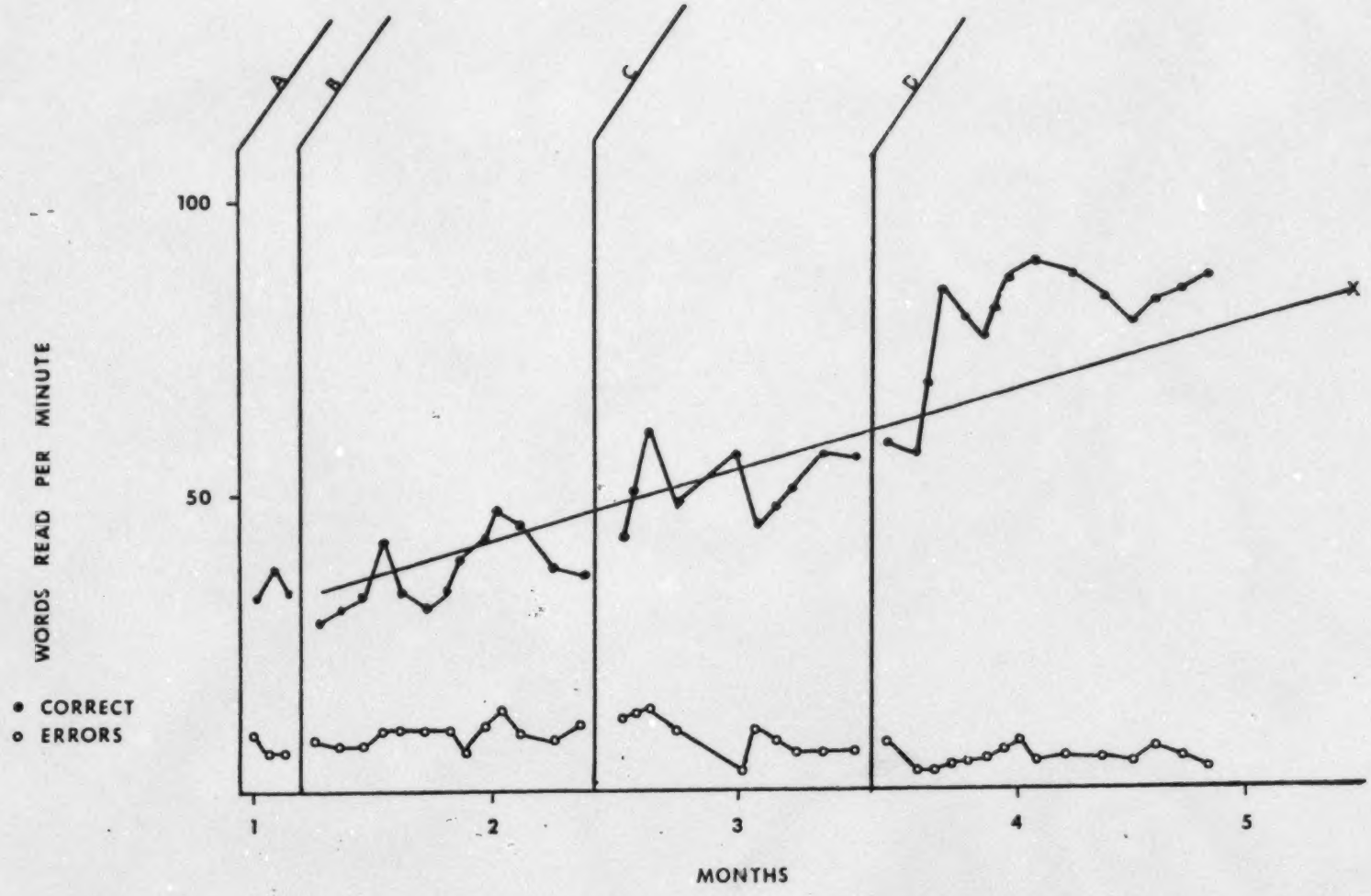


Figure 3. Number of Correct Words (•) and Errors (o) Per Minute Read by Michael from Pages in SRA, Level 2 Across Time, Under Baseline (A) and Three Instructional Strategies (B, C, and D).

Instructional Change Form

Instructional Procedures	Arrangement	Time	Materials	Motivational Strategies
Oral Reading Practice Comprehension exercises	Group (1:5)	45 minutes	Double Action Short Story, Part 2 Story Writing & class discussion	Generating own stories
Language Experience Approach	Individual with para-professional	same	Student's own stories File cards Story Folder	same
Language Experience	Individual with para-professional	20 minutes	See above	same
Reading Comprehension Activities	Individual with teacher	20 minutes	McCall-Crabbs, Book E SRA kit	individual arrangement with teacher

41

40

Figure 4. Michael's Instructional Change Form

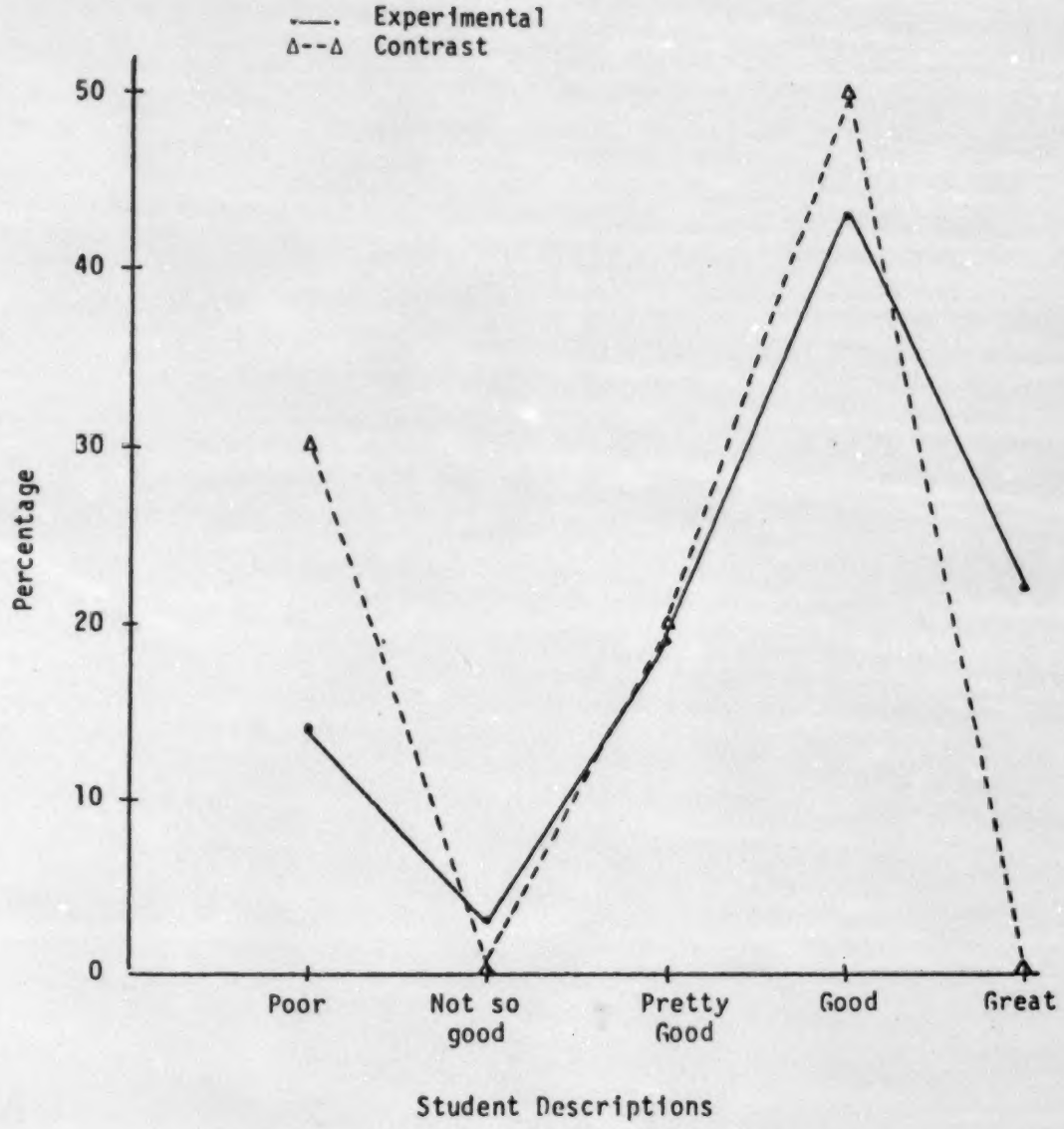


Figure 5. Percentage of Student Descriptions of Progress During the Current Academic Year.

Appendix A

Accuracy of Implementation Rating Scale

School:_____ Student:_____

Date:_____ Teacher:_____

Observer (Items 1 and 2):_____

Rater (Items 3-13):_____

Number of observations prior to rating:_____

Time observation begins:_____ Time observation ends:_____

Time allocated to reading instruction per day:_____

Curriculum used for measurement: Publisher_____

Series_____ Level_____

Instructions

Circle the number that accurately reflects your rating for each variable. Only one number may be circled per variable. 1 reflects a low level of implementation and 5 means total implementation of the Procedures to Develop and Monitor Progress on IEP Goals. See Operational Definitions. Items 1 and 2 require direct observation of the measurement administration. Items 3, 4, 5, 6, and 7 require inspection of the student graph. Items 8, 9, and 10 require inspection of the student's IEP form. The Instructional Plan must be inspected to rate item 11. The Change Record must be inspected to rate items 12 and 13.

Dir. Obs.	1. Administering the Measurement Task	1	2	3	4	5
	2. Selecting the Stimulus Material	1	2	3	4	5
Inspect Graph	3. Sampling for Instructional Level	1	2	3	4	5
	4. Baseline	1	2	3	4	5
	5. Graph Set-up	1	2	3	4	5
	6. Aimline	1	2	3	4	5
	7. Timing of Instructional Changes	1	2	3	4	5
Inspect IEP	8. Long-Range Goal	1	2	3	4	5
	9. Short-Term Objective	1	2	3	4	5
	10. Measurement System	1	2	3	4	5
Insp. Inst. Plan	11. Instructional Plan	1	2	3	4	5
Insp. Change Rec.	12. Substantial Changes	1	2	3	4	5
	13. One, Clear Change	1	2	3	4	5

AIRS

Operational Definitions

Accuracy of Implementation Rating Scale

1. Administering the Measurement Task

5 - The measurement task is administered correctly: teacher brings stopwatch and pencil to measurement area; gives correct directions for the task; administers the measurement procedure for one minute; correctly marks the teacher copy; correctly counts words correct and incorrect; correctly counts words correct and incorrect; correctly plots the data point.

1 - The teacher: forgets necessary materials; does not give directions; does not time the task accurately; fails to mark the teacher copy or incorrectly marks errors; miscounts correct and incorrect words; and inaccurately plots the data point.

2. Selecting the Stimulus Material

5 - The teacher has followed these procedures: Uses passages selected from the level that represents the annual goal. Observers should record the book from which the passage was selected and later check this with the long-range goal level. At this level find the pages in these stories that do not have excessive dialogue, indentations, and/or unusual pronouns. Write these page numbers on equal size slips of paper.

- Put the slips of paper into a drawbag and shake it.

- Randomly pick a slip of paper.

- The page number chosen is the page where the student begins reading. If the page chosen is a passage that was read earlier during the week, draw another page number.

Other completely random procedures are also rated a 5. If, however, not all passages have an equal chance of being selected, a 4 rating would be indicated.

1 - The teacher fails to randomly pick the passage or the sample is taken from a domain which is greater or smaller than the one indicated in the goal.

3. Sampling for Instructional Level

5 - The teacher has sampled from higher or lower reading levels to find the level in which the student reads 20-29 wpm (grades 1 & 2) or 30-39 wpm (grades 3 and up).

1 - The teacher is measuring at a level which is too high or too low.

4. Baseline

5 - The student's performance has been measured at least 3 times to establish a stable baseline. A stable baseline means that all data points fall within a range of 10.

1 - The teacher has not found a level for which a stable baseline has been established or has failed to collect 3 data points during the baseline phase.

5. Graph Set-Up

5 - The graph is accurately set up: The dates filled in on the horizontal axis; the vertical axis is correctly labeled words read per minute from _____ material; the units of measurement are specified; the student's name and subject area are certified; a key identifies the symbols for correct (.) and incorrect (x); symbols are placed at the intersection of date and score; the data points are connected with straight lines; and absences are recorded on the graph as (abs.).

1 - The graph does not include many of the items mentioned above.

6. Aimline

5 - The long-range goal is marked on the graph with an X at the intersection of the desired performance level and date of attainment and a line of desired progress connects the point representing the student's median score of the last 3 data points from baseline and the LRG.

1 - The long-range goal is not marked on the graph and/or the median and LRG are not connected.

7. Timing of Instructional Changes

5 - All the adjustments in the student's program are made at the appropriate time given the rules for data utilization:

(1) Compare the actual slope based on 7 to 10 data points to the slope required to attain the Annual Goal.

(2) If the actual slope is equal to, or steeper than, the Annual Goal slope, continue the program.

(3) If the actual slope is flatter than the Annual Goal slope, change the program.

1 - None of the adjustments in the student's program are made at the appropriate time.

8. Long-Range Goal

5 - The long-range goal is accurately written; goal specifies the number of weeks until next review; stimulus materials for the goal represents the level in which the student is performing at entry level criterion; goal specifies student behavior; goal specifies mastery criterion of 50 wpm with fewer than 5 errors (grades 1 & 2) or 70 wpm with fewer than 7 errors (grades 3-5) when there are 36 weeks until the annual review. If there are fewer than 36 weeks, the criteria can be lowered proportionately.

1 - The long-range goal contains none of the above criteria.

9. Short-Term Objective

5 - The short-term objective is accurately written; stimulus material and behavior is specified; and the average increase in performance is the desired performance minus the actual performance divided by the number of weeks until the annual review.

1 - The short-term objective contains none of the above criteria.

10. Measurement System

5 - The teacher has indicated how the material is organized, the frequency of measurement, and what is to be recorded on the graph.

1 - The measurement system is not specified.

11. Instructional Plan

5 - The instructional plan includes clear and specific descriptions of the instructional procedures, the time spent in each activity, the pertinent materials, the arrangements, and the motivational strategies.

1 - The instructional plan is unclear and lacks specific descriptions of the instructional procedures, the time spent in each activity, the pertinent materials, the arrangements, and the motivational strategies.

12. Substantial Changes

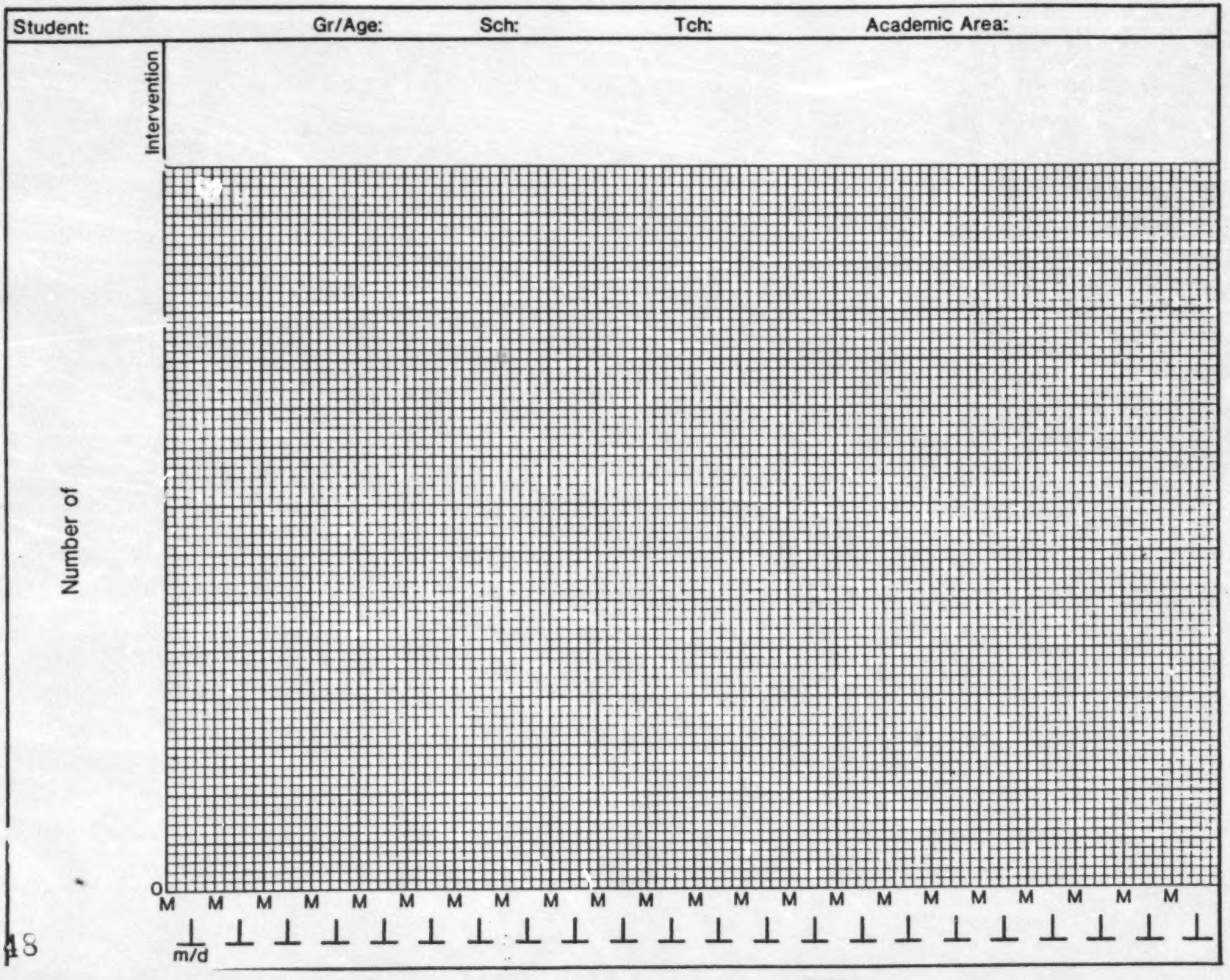
5 - The adjustments in the student's program are always substantial (have a good chance of being effective; see Unit XIV).

1 - The adjustments are never substantial.

13. Clear Change

5 - All the adjustments made introduce only one, clear program change.

1 - All the adjustments made introduce more than one change and/or the change is unclear.


Student:
Gr/Age:
Sch:
Tch:
Academic Area:
Intervention
Number of
0
M M M M M M M M M M M M M M M M M M M M M M
m/d

Appendix B

Appendix C

GOAL AND OBJECTIVE FORM

GOAL In _____ (# school weeks until year's end), when presented with stories from _____ (Level #, series), _____ (student's name) will read aloud at the rate of _____ (wpm correct) with no more than _____ (#) errors.

OBJECTIVE Each successive week, when presented with a random selection from _____ (same level # and series as above), student will read aloud at an average increase of _____ (#) wpm and no increase in errors.

MEASUREMENT SYSTEM Frequency: at least 2X/week; stimulus format: oral reading passages; Test administration procedure: same as manual; Scoring procedure: same as manual; charting conventions: same as manual.

Student Name _____ IEP Goal Area _____ Time Available for Instruction _____ Location of Instruction _____

Changes in the Instructional Plan

Instructional Procedures	Arrangement	Time	Materials	Motivational Strategies
DATE:				
DATE:				
DATE:				
DATE:				
DATE:				

Appendix E

END-OF-YEAR-STUDENT INTERVIEW

Student's Name_____ Date_____

Data Collector's Name_____

1) What kind of progress have you made in reading this year?

☐ Poor ☐ Not-so-Good ☐ Pretty Good ☐ Good ☐ Great ☐ Don't Know

2) Do you know what your goal is in reading for this year? ☐ Yes ☐ No

If yes, can you tell it to me. (Write below)

_____

_____

3) Will you meet your reading goal this year? ☐ Yes ☐ No

4) How do you know if you are going to meet your reading goal this year?

_____

_____

Student's Accuracy

Please indicate the accuracy of the student's responses by answering the following questions:

1) Did the student name correctly his/her reading goal? ☐ Yes ☐ No

If no, how was it wrong?_____

_____

_____

_____

2) Did the student indicate correctly whether he/she will meet his/her goal in reading?

☐ Yes ☐ No

If no, what was wrong?_____

_____

_____

Appendix F

Sample Experimental Training Schedule

Teacher Trainer_____ Date_____

Teacher _____

School _____

Below, please: (a) describe the nature of the training you have provided this teacher during the study, (b) indicate the administrative arrangement under which the training was provided (individual, 1-3 ratio, etc.), and (c) estimate the number of hours you spent with the teacher on each training area.

DESCRIPTION OF TRAINING	ARRANGEMENT	NUMBER OF HOURS
data-based monitoring procedures	individual	15 hrs.
meeting the individual needs of the students	individual	1 hr.
teaching decoding and comprehension (literal and inferential) skills	individual	3 hrs.
classroom management for reading	individual	1 hr.

Sample Control Training Schedule

Teacher Trainer _____ Date _____

Teacher _____

School _____

Below, please: (a) describe the nature of the training you have provided this teacher during the study, (b) indicate the administrative arrangement under which the training was provided (individual, 1-3 ratio, etc.), and (c) estimate the number of hours you spent with the teacher on each training area.

DESCRIPTION OF TRAINING	ARRANGEMENT	NUMBER OF HOURS
Using the language experience approach	individual	10 hrs.
Individualizing	individual	5 hrs
Using audio-visual equipment for teaching and motivating	individual	2 hrs.
Diagnostic assessment tools	individual	3 hrs.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $3.00 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor, IRLD, 350 Elliott Hall; 75 East River Road, University of Minnesota, Minneapolis, MN 55455.

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota round-table conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Note: Monographs No. 1 - 6 and Research Report No. 2 are not available for distribution. These documents were part of the Institute's 1979-1980 continuation proposal, and/or are out of print.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

Mirkin, P. K., & Deno, S. L. Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). August, 1979.

Thurlow, M. L., & Ysseldyke, J. E. Current assessment and decision-making practices in model programs for the learning disabled (Research Report No. 11). August, 1979.

Deno, S. L., Chiang, B., Tindal, G., & Blackburn, M. Experimental analysis of program components: An approach to research in CSDC's (Research Report No. 12). August, 1979.

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. Similarities and differences between underachievers and students labeled learning disabled: Identical twins with different mothers (Research Report No. 13). September, 1979.

Ysseldyke, J., & Algozzine, R. Perspectives on assessment of learning disabled students (Monograph No. 11). October, 1979.

Poland, S. F., Ysseldyke, J. E., Thurlow, M. L., & Mirkin, P. K. Current assessment and decision-making practices in school settings as reported by directors of special education (Research Report No. 14). November, 1979.

McGue, M., Shinn, M., & Ysseldyke, J. Validity of the Woodcock-Johnson psycho-educational battery with learning disabled students (Research Report No. 15). November, 1979.

Deno, S., Mirkin, P., & Shinn, M. Behavioral perspectives on the assessment of learning disabled children (Monograph No. 12). November, 1979.

Sutherland, J. H., Algozzine, B., Ysseldyke, J. E., & Young, S. What can I say after I say LD? (Research Report No. 16). December, 1979.

Deno, S. L., & Mirkin, P. K. Data-based IEP development: An approach to substantive compliance (Monograph No. 13). December, 1979.

Ysseldyke, J., Algozzine, B., Regan, R., & McGue, M. The influence of test scores and naturally-occurring pupil characteristics on psycho-educational decision making with children (Research Report No. 17). December, 1979.

Algozzine, B., & Ysseldyke, J. E. Decision makers' prediction of students' academic difficulties as a function of referral information (Research Report No. 18). December, 1979.

Ysseldyke, J. E., & Algozzine, B. Diagnostic classification decisions as a function of referral information (Research Report No. 19). January, 1980.

Deno, S. L., Mirkin, P. K., Chiang, B., & Lowry, L. Relationships among simple measures of reading and performance on standardized achievement tests (Research Report No. 20). January, 1980.

Deno, S. L., Mirkin, P. K., Lowry, L., & Kuehnle, K. Relationships among simple measures of spelling and performance on standardized achievement tests (Research Report No. 21). January, 1980.

Deno, S. L., Mirkin, P. K., & Marston, D. Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). January, 1980.

Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehnle, K. Formative evaluation: Continued development of data utilization systems (Research Report No. 23). January, 1980.

Deno, S. L., Mirkin, P. K., Robinson, S., & Evans, P. Relationships among classroom observations of social adjustment and sociometric rating scales (Research Report No. 24). January, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Factors influential on the psychoeducational decisions reached by teams of educators (Research Report No. 25). February, 1980.

Ysseldyke, J. E., & Algozzine, B. Diagnostic decision making in individuals susceptible to biasing information presented in the referral case folder (Research Report No. 26). March, 1980.

Thurlow, M. L., & Greener, J. W. Preliminary evidence on information considered useful in instructional planning (Research Report No. 27). March, 1980.

Ysseldyke, J. E., Regan, R. R., & Schwartz, S. Z. The use of technically adequate tests in psychoeducational decision making (Research Report No. 28). April, 1980.

Richey, L., Potter, M., & Ysseldyke, J. Teachers' expectations for the siblings of learning disabled and non-learning disabled students: A pilot study (Research Report No. 29). May, 1980.

Thurlow, M. L., & Ysseldyke, J. E. Instructional planning: Information collected by school psychologists vs. information considered useful by teachers (Research Report No. 30). June, 1980.

Algozzine, B., Webber, J., Campbell, M., Moore, S., & Gilliam, J. Classroom decision making as a function of diagnostic labels and perceived competence (Research Report No. 31). June, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., Richey, L., & Thurlow, M. L. Psychoeducational assessment and decision making: A computer-simulated investigation (Research Report No. 32). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R. R., Potter, M., & Richey, L. Psychoeducational assessment and decision making: Individual case studies (Research Report No. 33). July, 1980.

Ysseldyke, J. E., Algozzine, B., Regan, R., Potter, M., & Richey, L. Technical supplement for computer-simulated investigations of the psychoeducational assessment and decision-making process (Research Report No. 34). July, 1980.

Algozzine, B., Stevens, L., Costello, C., Beattie, J., & Schmid, R. Classroom perspectives of LD and other special education teachers (Research Report No. 35). July, 1980.

Algozzine, B., Siders, J., Siders, J., & Beattie, J. Using assessment information to plan reading instructional programs: Error analysis and word attack skills (Monograph No. 14). July, 1980.

Ysseldyke, J., Shinn, M., & Epps, S. A comparison of the WISC-R and the Woodcock-Johnson Tests of Cognitive Ability (Research Report No. 36). July, 1980.

Algozzine, B., & Ysseldyke, J. E. An analysis of difference score reliabilities on three measures with a sample of low achieving youngsters (Research Report No. 37). August, 1980.

Shinn, M., Algozzine, B., Marston, D., & Ysseldyke, J. A theoretical analysis of the performance of learning disabled students on the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 38). August, 1980.

Richey, L. S., Ysseldyke, J., Potter, M., Regan, R. R., & Greener, J. Teachers' attitudes and expectations for siblings of learning disabled children (Research Report No. 39). August, 1980.

Ysseldyke, J. E., Algozzine, B., & Thurlow, M. L. (Eds.). A naturalistic investigation of special education team meetings (Research Report No. 40). August, 1980.

Meyers, B., Meyers, J., & Deno, S. Formative evaluation and teacher decision making: A follow-up investigation (Research Report No. 41). September, 1980.

Fuchs, D., Garwick, D. R., Featherstone, N., & Fuchs, L. S. On the determinants and prediction of handicapped children's differential test performance with familiar and unfamiliar examiners (Research Report No. 42). September, 1980.

Algozzine, B., & Stoller, L. Effects of labels and competence on teachers' attributions for a student (Research Report No. 43). September, 1980.

Ysseldyke, J. E., & Thurlow, M. L. (Eds.). The special education assessment and decision-making process: Seven case studies (Research Report No. 44). September, 1980.

Ysseldyke, J. E., Algozzine, B., Potter, M., & Regan, R. A descriptive study of students enrolled in a program for the severely learning disabled (Research Report No. 45). September, 1980.

Marston, D. Analysis of subtest scatter on the tests of cognitive ability from the Woodcock-Johnson Psycho-Educational Battery (Research Report No. 46). October, 1980.

Algozzine, B., Ysseldyke, J. E., & Shinn, M. Identifying children with learning disabilities: When is a discrepancy severe? (Research Report No. 47). November, 1980.

Fuchs, L., Tindal, J., & Deno, S. Effects of varying item domain and sample duration on technical characteristics of daily measures in reading (Research Report No. 48). January, 1981.

Marston, D., Lowry, L., Deno, S., & Mirkin, P. An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study (Research Report No. 49). January, 1981.

Marston, D., & Deno, S. The reliability of simple, direct measures of written expression (Research Report No. 50). January, 1981.

Epps, S., McGue, M., & Ysseldyke, J. E. Inter-judge agreement in classifying students as learning disabled (Research Report No. 51). February, 1981.

Epps, S., Ysseldyke, J. E., & McGue, M. Differentiating LD and non-LD students: "I know one when I see one" (Research Report No. 52). March, 1981.

Evans, P. R., & Peham, M. A. S. Testing and measurement in occupational therapy. A review of current practice with special emphasis on the Southern California Sensory Integration Tests (Monograph No. 15). April, 1981.

Fuchs, L., Wesson, C., Tindal, G., & Mirkin, P. Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). June, 1981.

Fuchs, D., Featherstone, N., Garwick, D. R., & Fuchs, L. S. The importance of situational factors and task demands to handicapped children's test performance (Research Report No. 54). June, 1981.

Tindal, G., & Deno, S. L. Daily measurement of reading: Effects of varying the size of the item pool (Research Report No. 55). July, 1981.

Fuchs, L. S., & Deno, S. L. A comparison of teacher judgment, standardized tests, and curriculum-based approaches to reading placement (Research Report No. 56). August, 1981.

Fuchs, L., & Deno, S. The relationship between curriculum-based mastery measures and standardized achievement tests in reading (Research Report No. 57). August, 1981.

Christenson, S., Graden, J., Potter, M., & Ysseldyke, J. Current research on psychoeducational assessment and decision making: Implications for training and practice (Monograph No. 16). September, 1981.

Christenson, S., Ysseldyke, J., & Algozzine, B. Institutional constraints and external pressures influencing referral decisions (Research Report No. 58). October, 1981.

Fuchs, L., Fuchs, D., & Deno, S. Reliability and validity of curriculum-based informal reading inventories (Research Report No. 59). October, 1981.

Algozzine, B., Christenson, S., & Ysseldyke, J. Probabilities associated with the referral-to-placement process (Research Report No. 60). November, 1981.

Tindal, G., Fuchs, L. Christenson, S., Mirkin, P., & Deno, S. The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). November, 1981.

Mirkin, P., Fuchs, L., Tindal, G., Christenson, S., & Deno, S. The effect of IEP monitoring strategies on teacher behavior (Research Report No. 62). December, 1981.

Wesson, C., Mirkin, P., & Deno, S. Teachers' use of self instructional materials for learning procedures for developing and monitoring progress on IEP goals (Research Report No. 63). January, 1982.

Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). January, 1982.

Potter, M., & Mirkin, P. Instructional planning and implementation practices of elementary and secondary resource room teachers: Is there a difference? (Research Report No. 65). January, 1982.

Thurlow, M. L., & Ysseldyke, J. E. Teachers' beliefs about LD students (Research Report No. 66). January, 1982.

Graden, J., Thurlow, M. L., & Ysseldyke, J. E. Academic engaged time and its relationship to learning: A review of the literature (Monograph No. 17). January, 1982.

King, R., Wesson, C., & Deno, S. Direct and frequent measurement of student performance: Does it take too much time? (Research Report No. 67). February, 1982.

Greener, J. W., & Thurlow, M. L. Teacher opinions about professional education training programs (Research Report No. 68). March, 1982.

Algozzine, B., & Ysseldyke, J. Learning disabilities as a subset of school failure: The oversophistication of a concept (Research Report No. 69). March, 1982.

Fuchs, D., Zern, D. S., & Fuchs, L. S. A microanalysis of participant behavior in familiar and unfamiliar test conditions (Research Report No. 70). March, 1982.

Shinn, M. R., Ysseldyke, J., Deno, S., & Tindal, G. A comparison of psychometric and functional differences between students labeled learning disabled and low achieving (Research Report No. 71). March, 1982.

Thurlow, M. L. Graden, J., Greener, J. W., & Ysseldyke, J. E. Academic responding time for LD and non-LD students (Research Report No. 72). April, 1982.

Graden, J., Thurlow, M., & Ysseldyke, J. Instructional ecology and academic responding time for students at three levels of teacher-perceived behavioral competence (Research Report No. 73). April, 1982.

Algozzine, B., Ysseldyke, J., & Christenson, S. The influence of teachers' tolerances for specific kinds of behaviors on their ratings of a third grade student (Research Report No. 74). April, 1982.

Wesson, C., Deno, S., & Mirkin, P. Research on developing and monitoring progress on IEP goals: Current findings and implications for practice (Monograph No. 18). April, 1982.

Mirkin, P., Marston, D., & Deno, S. L. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students (Research Report No. 75). May, 1982.

Algozzine, B., Ysseldyke, J., Christenson, S., & Thurlow, M. Teachers' intervention choices for children exhibiting different behaviors in school (Research Report No. 76). June, 1982.

Tucker, J., Stevens, L. J., & Ysseldyke, J. E. Learning disabilities: The experts speak out (Research Report No. 77). June, 1982.

Thurlow, M. L., Ysseldyke, J. E., Graden, J., Greener, J. W., & Mecklenberg, C. Academic responding time for LD students receiving different levels of special education services (Research Report No. 78). June, 1982.

Graden, J. L., Thurlow, M. L., Ysseldyke, J. E., & Algozzine, B. Instructional ecology and academic responding time for students in different reading groups (Research Report No. 79). July, 1982.

Mirkin, P. K., & Potter, M. L. A survey of program planning and implementation practices of LD teachers (Research Report No. 80). July, 1982.

Fuchs, L. S., Fuchs, D., & Warren, L. M. Special education practice in evaluating student progress toward goals (Research Report No. 81). July, 1982.

Kuehnle, K., Deno, S. L., & Mirkin, P. K. Behavioral measurement of social adjustment: What behaviors? What setting? (Research Report No. 82). July, 1982.

Fuchs, D., Dailey, Ann Madsen, & Fuchs, L. S. Examiner familiarity and the relation between qualitative and quantitative indices of expressive language (Research Report No. 83). July, 1982.

Videen, J., Deno, S., & Marston, D. Correct word sequences: A valid indicator of proficiency in written expression (Research Report No. 84). July, 1982.

Potter, M. L. Application of a decision theory model to eligibility and classification decisions in special education (Research Report No. 85). July, 1982.

Greener, J. E., Thurlow, M. L., Graden, J. L., & Ysseldyke, J. E. The educational environment and students' responding times as a function of students' teacher-perceived academic competence (Research Report No. 86). August, 1982.

Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study (Research Report No. 87). August, 1982.

Skiba, R., Wesson, C., & Deno, S. L. The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 88). September, 1982.

Martson, D., Tindal, G., & Deno, S. L. Eligibility for learning disability services: A direct and repeated measurement approach (Research Report No. 89). September, 1982.

Thurlow, M. L., Ysseldyke, J. E., & Graden, J. L. LD students' active academic responding in regular and resource classrooms (Research Report No. 90). September, 1982.

Ysseldyke, J. E., Christenson, S., Pianta, R., Thurlow, M. L., & Algozzine, B. An analysis of current practice in referring students for psychoeducational evaluation: Implications for change (Research Report No. 91). October, 1982.

Ysseldyke, J. E., Algozzine, B., & Epps, S. A logical and empirical analysis of current practices in classifying students as handicapped (Research Report No. 92). October, 1982.

Tindal, G., Marston, D., Deno, S. L., & Germann, G. Curriculum differences in direct repeated measures of reading (Research Report No. 93). October, 1982.

Fuchs, L.S., Deno, S. L., & Marston, D. Use of aggregation to improve the reliability of simple direct measures of academic performance (Research Report No. 94). October, 1982.

Ysseldyke, J. E., Thurlow, M. L., Mecklenburg, C., & Graden, J. Observed changes in instruction and student responding as a function of referral and special education placement (Research Report No. 95). October, 1982.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. Effects of frequent curriculum-based measurement and evaluation on student achievement and knowledge of performance: An experimental study (Research Report No. 96). November, 1982.